

# Investigating the role of musical genre in human perception of music stretching resistance \*

Jun Chen<sup>+,1</sup> and Chaokun Wang<sup>\*,1</sup>

<sup>1</sup>School of Software, Tsinghua University, Beijing, 100084, P.R. China.

<sup>+</sup>chenjun14@mails.thu.edu.cn

<sup>\*</sup>chaokun@tsinghua.edu.cn

## ABSTRACT

To stretch a music piece to a given length is a common demand in people's daily lives, e.g., in audio-video synchronization and animation production. However, it is not always guaranteed that the stretched music piece is acceptable for general audience since music stretching suffers from people's perceptual artefacts. Over-stretching a music piece will make it uncomfortable for human psychoacoustic hearing. The research on music stretching resistance attempts to estimate the maximum stretchability of music pieces to further avoid over-stretch. It has been observed that musical genres can significantly improve the accuracy of automatic estimation of music stretching resistance, but how musical genres are related to music stretching resistance has never been explained or studied in detail in the literature. In this paper, the characteristics of music stretching resistance are compared across different musical genres. It is found that music stretching resistance has strong intra-genre cohesiveness and inter-genre discrepancies in the experiments. Moreover, the ambiguity and the symmetry of music stretching resistance are also observed in the experimental analysis. These findings lead to a new measurement on the similarity between different musical genres based on their music stretching resistance. In addition, the analysis of variance (ANOVA) also supports the findings in this paper by verifying the significance of musical genre in shaping music stretching resistance.

## Introduction

Music stretching resistance (*abbr.* MSR) which describes the acceptable range of time stretching rate of music piece for people's psychoacoustic hearing,<sup>1,2</sup> consists of the minimum compressing rate, denoted as  $\alpha_{min}$  ( $0.0 < \alpha_{min} < 1.0$ ), and the maximum elongating rate, denoted as  $\alpha_{max}$  ( $1.0 < \alpha_{max} < 2.0$ ).<sup>3</sup> The research into MSR is of broad interest in fields like time-scale modification of speech,<sup>1,2</sup> music resizing,<sup>4-6</sup> dynamic music re-scaling<sup>7</sup> as well as other fields related to human perception of music and psychoacoustic hearing. The computational method<sup>3</sup> which estimates MSR by incorporating sound features (e.g. spectral analysis, timbre, pitch) and musical genres, has shown that musical genre is much more important in affecting MSR compared with sound features like timbre, pitch and rhythm. However, there is still no in-depth research to investigate the relationship between MSR and musical genres to the best of our knowledge.

Generally speaking, the existence of MSR can be attributed to the process of receiving and recognizing accelerated and decelerated sounds where people would make positive or negative reaction to accept/reject the changes based on the satisfaction of their psychoacoustic hearing. Basically, MSR is related to the perception of music and the artefacts of digital signal processing. Human perception of music has much in common with human recognition of natural languages, where the structures (syntax/harmony), the vocabularies (words/chords), the tonal properties (inflection/timbre) and the temporal features (prosody/rhythm) are shared.<sup>8</sup> Meanwhile, the functional magnetic resonance imaging (fMRI) shows greater neuronal activity in the voice-selective regions of participants when they are listening to vocal sounds than to non-vocal sounds,<sup>9</sup> which indicates that people may also have different degrees of sensitivity towards music with/without lyrics, or with sparse/dense lyrics. This difference may contribute to the creation of MSR since stretching operations change the density of lyrics (e.g. words per minute), too. The compressing or elongating operations<sup>1,2,4-6</sup> on a given music piece change the tempos while preserving the pitch features,<sup>1,2,10</sup> which leads to a range of acceptability on user preferred speeds<sup>11</sup> as well as user tolerable speed ranges. These are the evidences why MSR exists for general audience.

The dynamic attending theories<sup>12-15</sup> posit that the internal mechanism, like a clock, a time keeper or an oscillator inside human beings, resonates and synchronizes with the periodicity of stretched music pieces to enable the audience to follow changes in tempo. Without the knowledge of MSR, however, perceptual artefacts<sup>6</sup> are more likely to occur and degrade the auditory experience of general listeners when a music piece is stretched at a rate out of its acceptable stretching range, i.e.,

\*If you are interested in the collected data, please contact Dr. Chen: chenjun14@mails.thu.edu.cn.

overly compressed or overly elongated. Most listeners are likely to identify the base tempo of music around 120 BPM (beats per minute), and the acceleration or the deceleration in speed usually induces more ambiguity in the identification of tempo.<sup>16,17</sup> To some extent, the naïve approximation of acceptable stretching range, i.e.,  $\pm 20\%$  (namely  $\alpha_{min} = 0.80$ ,  $\alpha_{max} = 1.20$ ),<sup>18</sup> is baseless and fails to utilize the features of music pieces, considering that the perception of tempo appropriateness for a given music piece is determined by its contents.<sup>19</sup> One of the categorical features of music is genre which incorporates cultural backgrounds and emotions of artists, and characterizes the similarity between music pieces.<sup>20</sup> Meanwhile, musical genres are also related to music content features, e.g., timbre, pitch and rhythm, so that the automatic genre classification algorithms<sup>20–23</sup> could work, which makes genre a potential factor in studying MSR.

In Chen's previous work,<sup>3</sup> MSR values are discretized as labels along the axis of stretching rate. The estimation of MSR is performed by classifying the label of the given music piece using its sound features (spectral analysis, timbre, pitch and tempo) and musical genre with the machine learning techniques. It is observed that musical genre has larger contribution to improve the classification accuracy compared with sound features. But how musical genre is related to MSR has never been studied in detail. We believe that it is necessary to further explore the relationship between MSR and musical genres.

In this paper, we investigate the important role of musical genres in shaping MSR. We find that MSR tends to be constant with small fluctuation within a given musical genre. The significance of musical genres in the analysis of variance substantiates the existence of inter-genre discrepancy and intra-genre cohesiveness in MSR, i.e., MSR values are widely discrepant among different genres and are inherently cohesive within a same genre. The ambiguity of MSR is also diversified for different musical genres. Besides, the MSR values of a given musical genre are symmetric in terms of the range boundaries (i.e., mean) as well as the ambiguity of range boundaries (i.e., standard deviation). The regression lines from original tempos of music pieces to MSR values ( $\alpha_{min}$  and  $\alpha_{max}$ ) are almost horizontal within a given musical genre in the experiments, which further indicates that MSR has little correlation with tempo unlike that with musical genre. Besides, we also analyze the MSR-based similarity between musical genres by computing the overlapping area of covering regions on the  $\alpha_{min}$ – $\alpha_{max}$  coordinate system. This new measurement of similarity offers a new perspective on musical genres based on human perception of music stretching resistance. MSR is a psychoacoustic reflection of human perception of music,<sup>11</sup> and the study on MSR also sheds new light on content-aware music adaption<sup>4–6</sup> and dynamic attending theory.<sup>12–15</sup>

## Methods

### Participants

We recruited 17 college students as participants in the experiments. These participants ranged in age from 18 to 25 years old, with 29.4% female and 70.6% male. Most of them were non-musicians except for two participants who had received piano education for more than 3 years by the time of experiments. It has been proved that listeners (musicians or non-musicians) can make consistent judgments on whether music pieces are played overly fast or overly slow.<sup>8,19</sup> The composition of participants in our experiments is similar to the real situation in our daily lives. The participants were selected among the people who enjoyed listening to music and were willing to spend at least half an hour every day to conduct the experiments. The experiments last for about one month so that the experimental results would be less influenced by the short-term changes of participants' physical or mental states, such as moods (happy, sad), time (morning, evening), locations (home, workplace), and weather (rainy, sunny). Since the participants were expected to conduct the listening experiments for a long period of time, we excluded those short-term participants who could not make through the one-month experiments to get the results from the final 17 participants. Although the final number of participants is not very large, we tried to minimize the impact of personal preferences by increasing the overlap of music pieces that different participants listened, and majority-voting the results they reported. All the participants were paid a little bit for their endeavor when the experiments finished.

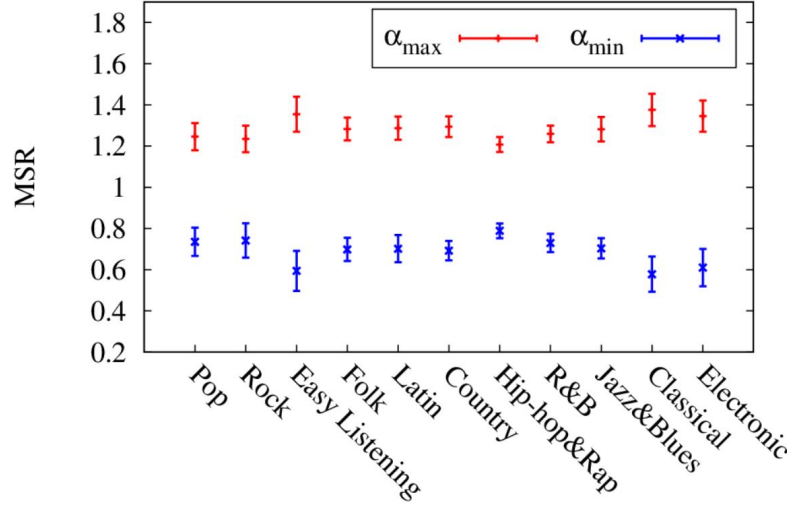
### Experimental Settings

We used the collection of 894 songs from<sup>3</sup> in this study. These songs identically cover 11 musical genres as shown in Table 1. These songs were randomly crawled from a music website<sup>1</sup>, and the genres of these songs were annotated by referring to the meta data of these songs as well as the genre taxonomy from Wikipedia and some popular music websites.

All the songs in the collection were stretched in time domain using the synchronized overlap-add (SOLA) method<sup>24</sup> to avoid pitch shift, which was implemented in the SoundTouch library<sup>2</sup>. Each song was stretched into different versions of discrete stretching rates between (0.00, 2.00) with a rate step 0.02. That is, each song has 49 compressed versions with stretching rates in  $\{0.98, 0.96, \dots, 0.02\}$  as well as 49 elongated versions with stretching rates in  $\{1.02, 1.04, \dots, 1.98\}$ . Please note that no music piece with stretching rate beyond 2.00 will be acceptable from our experience since the elongation would have destroyed the musical structure too much and made the elongated music piece sound uncomfortable.

<sup>1</sup><http://www.top100.cn>

<sup>2</sup><http://www.surina.net/soundtouch>



**Figure 1.** The error bars of MSR distribution of 11 different musical genres.

Although there are some other different music stretching methods,<sup>4-7</sup> we choose SOLA<sup>24</sup> in this work because SOLA is a more fundamental work in the music stretching literature and it uniformly stretches music pieces. Therefore, the results reported in this paper are about uniformly music stretching.

### Procedure

Each participant was delivered at least one package of 20 random songs from our collection as well as the 98 stretched versions (49 compressed + 49 elongated) for each song. The genre composition in each song package is random, and some participants were delivered more song packages since they were faster in conducting the listening experiments. The participants were asked to listen to these stretched versions one after another, and judge whether each stretched version is acceptable. They could choose the order of stretching rates to listen in their preferred ways as long as they could give judgement on  $\alpha_{min}$  and  $\alpha_{max}$ . We offered the participants a few general judging criteria, e.g., speed acceptance, lyrics density acceptance and overall listening acceptance. The degradation of sound quality after stretch using SOLA method also turns out to be an important factor to influence people's psychoacoustic acceptance. For example, over elongation will make the sound interruptive while over compression will mess up the audio signals and make it sound noisy. Besides the given judging criteria, it is up to the participants to make their own judgement based on the comfort they feel in the listening. Though the total number of stretched versions of each song was not small, the participant did not have to listen to all of them, and they only needed to locate the minimum and the maximum acceptable stretching rates of each song in their package(s). To further help the participants to judge the acceptance of the stretched versions, we also developed a new music player for them, which supported 'one-click' switch between two stretched versions of a same song at the same position, e.g., 30% of the song.<sup>25</sup> With this music player, it would be easier for the participants to judge whether the listening version is over-stretched at the currently playing segment compared with the original version or the other acceptable versions.

When the participants completed the tasks in the delivered package(s), (s)he input the results, i.e., the  $\alpha_{min}$  and the  $\alpha_{max}$  of each song in the package, into our experimental web pages. Specifically, the participants enter the values of  $\alpha_{min}$  (e.g. 0.82) and  $\alpha_{max}$  (e.g. 1.26) of the songs in their package(s) through text fields on the web page. The participants were allowed to correct their results before the final submission by the end of the one-month experiment. Thus, the minimum compressing rate  $\alpha_{min}$  and the maximum elongating rate  $\alpha_{max}$  of each song in our collection were obtained for later analysis.

## Results and Discussion

We performed the analysis of variance (ANOVA) on the MSR values obtained from the listening experiments. MSR values represent  $\alpha_{min}$  and  $\alpha_{max}$  of each song reported by the participants. The results show that MSR values are significantly affected by musical genres (F-test, alpha levels are the 11 genres,  $\alpha_{min}$ :  $F(10, 883) = 74.683$ ,  $P < 0.001$ ,  $\alpha_{max}$ :  $F(10, 883) = 56.407$ ,  $P < 0.001$ ). Next, the relationship between MSR and musical genres is discussed in detail based on the experimental results.

### Basic MSR Properties

Fig. 1 illustrates the error bars (means and standard deviations) of MSR values for the 11 musical genres (the statistics are shown in Table 1), from which we can draw the following four basic conclusions about MSR:

Genre	#.Piece	$\alpha_{min}$	$\alpha_{max}$	Slope $\alpha_{min}$	Slope $\alpha_{max}$
Pop	87	$0.735 \pm 0.069$	$1.246 \pm 0.066$	$2.78 \times 10^{-4}$	$-5.30 \times 10^{-5}$
Rock	95	$0.742 \pm 0.083$	$1.235 \pm 0.064$	$6.85 \times 10^{-6}$	$-1.45 \times 10^{-4}$
Easy Listening	83	$0.594 \pm 0.097$	$1.355 \pm 0.086$	$-3.15 \times 10^{-4}$	$6.75 \times 10^{-5}$
Folk	80	$0.698 \pm 0.057$	$1.283 \pm 0.056$	$-4.29 \times 10^{-5}$	$-1.47 \times 10^{-4}$
Latin	68	$0.702 \pm 0.066$	$1.287 \pm 0.056$	$-1.34 \times 10^{-4}$	$-1.25 \times 10^{-5}$
Country	86	$0.692 \pm 0.047$	$1.294 \pm 0.050$	$1.73 \times 10^{-6}$	$1.79 \times 10^{-4}$
Hip-hop&Rap	82	$0.789 \pm 0.036$	$1.207 \pm 0.036$	$6.98 \times 10^{-5}$	$3.12 \times 10^{-6}$
R&B	90	$0.729 \pm 0.045$	$1.259 \pm 0.040$	$-2.77 \times 10^{-4}$	$-1.71 \times 10^{-5}$
Jazz&Blues	78	$0.703 \pm 0.049$	$1.282 \pm 0.059$	$-4.11 \times 10^{-5}$	$-1.24 \times 10^{-4}$
Classical	73	$0.578 \pm 0.085$	$1.376 \pm 0.079$	$-4.50 \times 10^{-4}$	$4.89 \times 10^{-4}$
Electronic	72	$0.610 \pm 0.090$	$1.346 \pm 0.076$	$-7.79 \times 10^{-5}$	$5.31 \times 10^{-5}$

**Table 1.** The means and deviations of the MSR values of different musical genres, as well as the slopes of regression lines from the original tempo of a music piece to  $\alpha_{min}/\alpha_{max}$  within each musical genre.

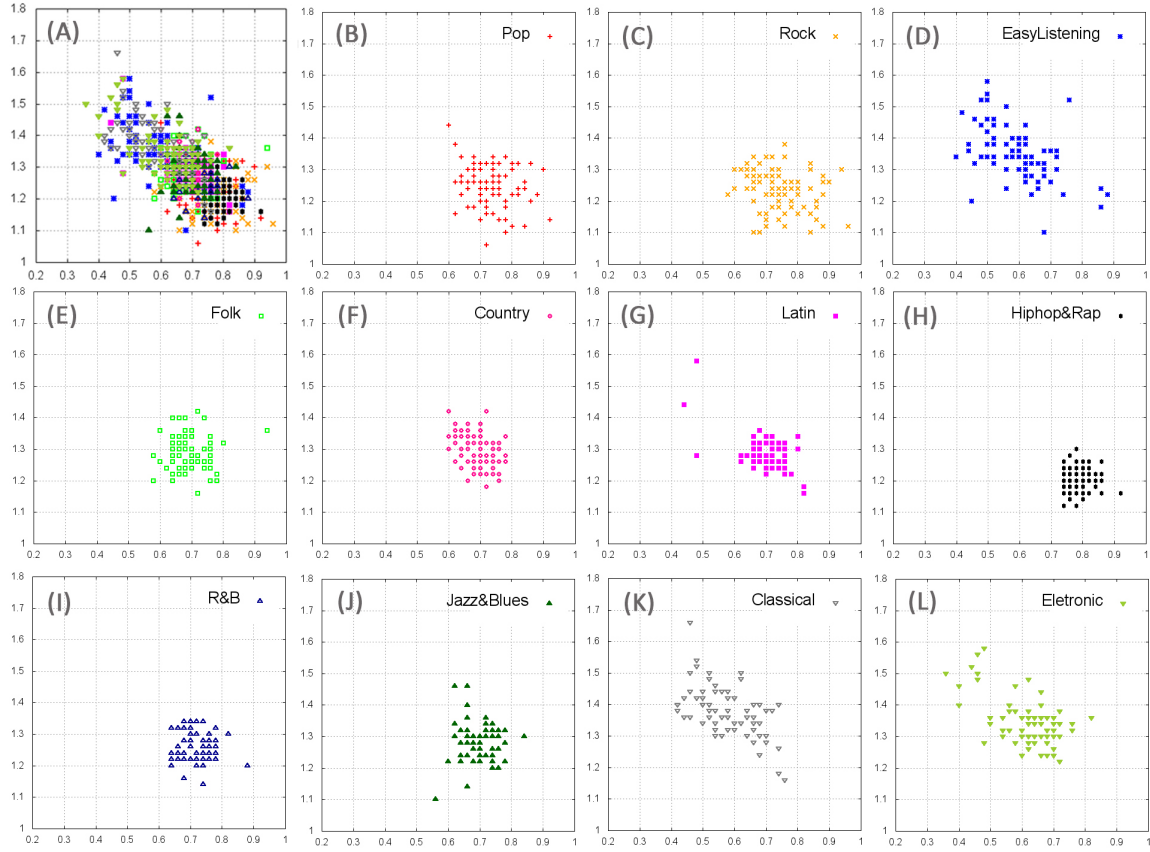
- **Inter-Genre Discrepancy:** Wide discrepancies in MSR values under different genres are observed. The position and the stretch of MSR of musical genres are quite different. For instance, *Easy Listening*, *Electronic* and *Classical* music have a wider stretching range (the interval between the mean of  $\alpha_{min}$  and that of  $\alpha_{max}$ ), while *Hip-hop&Rap* and *R&B* exhibit a narrower stretching reach. It is in line with the character effect (fast and slow) of music pieces on human preferred speeds and ranges of acceptability.<sup>11</sup>
- **Intra-Genre Cohesiveness:** MSR values are substantiated to be cohesive under a given genre according to the significance of musical genres in the analysis of variance.
- **Ambiguity:** The ambiguity of MSR is greater for *Easy Listening*, *Electronic* and *Classical* music pieces seen from the standard deviations. The larger the standard deviation is, the larger ambiguity the MSR of a musical genre is. This probably results from the rhythmic features of these musical genres since there may hardly be any fixed rhythmic patterns for these aforementioned genres, for instance, a piece of piano music or violin music. On the contrary, songs like *R&B*, *Hip-hop&Rap* usually follow a solid tempo throughout the whole music piece.
- **Symmetry:** Under a given genre, MSR tends to be symmetric between  $\alpha_{min}$  and  $\alpha_{max}$ , on both the range boundaries (mean) and the ambiguity of range boundaries (standard deviation). This property comes from the symmetric criteria that listeners used to judge  $\alpha_{min}$  and  $\alpha_{max}$ . Supposedly, the optimal tempo of a given song should occur near its base one. As a result, if a song is over-compressed and sounds uncomfortable, it is more likely that the elongated one with the same shift of stretching rate increase will also sound uneasy, and vice versa.

### Intra-Genre Linear Regression With Tempo

The stretching operations on a given music piece will lead to an inversely proportional relationship,  $t_s = \frac{t_o}{r}$ , where  $r$  is the stretching rate between the range (0.00, 2.00),  $t_o$  and  $t_s$  are the tempos of the original music piece and that of the stretched version measured by beat per minute, respectively. Evidently,  $t_o$  is fixed for a given music piece. Thus,  $t_s$  meets the upper limit when  $r$  equals  $\alpha_{min}$ , while  $t_s$  reaches the floor boundary when  $r$  equals  $\alpha_{max}$ . So as to identify the relationship between boundary tempos and MSR, the linear regression is performed from base tempos to  $\alpha_{min}$  and  $\alpha_{max}$  in the music collection, respectively. This is to study whether or not songs with different tempos under a given musical genre would generally have different  $\alpha_{min}$  and  $\alpha_{max}$ . The regression lines are almost horizontal under all genres since their slopes are very close to zero (Table 1). For example, the steepest slope in Table 1 is  $-4.89 \times 10^{-4}$  of  $\alpha_{max}$  of *Classical Music*. Since the original tempo of a music piece mostly varies between [0, 200] BPM, which can only cause less than 0.1 bias in the  $\alpha_{max}$  from other music pieces of this genre. Consequently, for a given musical genre, music pieces with different base tempos usually have similar MSR, which verifies the intra-genre cohesiveness of MSR from another point of view. Under a given genre, the limits of  $r$  are fixed, and thus the upper and the floor boundaries of  $t_s$  vary according to the value of  $t_o$  of a given music piece. The fact that the regressed slopes within each musical genre are almost zero is also a solid proof that MSR has little correlation with tempo unlike that with musical genre.

### MSR-based Musical Genre Similarity

To further investigate the relationship between MSR and musical genres, we illustrate the scatter plots of the MSR values within each musical genre in Fig. 2. Each point in these panels represents a pair ( $\alpha_{min}, \alpha_{max}$ ), while each point may correspond to a



**Figure 2.** The scatter plots of the MSR value distribution within each musical genre. The abscissa and the ordinate of each point in the panels represent the values of  $\alpha_{min}$  and  $\alpha_{max}$ , respectively. (A) is the assembly of the points from (B) to (L).

few songs sharing the same MSR values. Obviously, the covered region of the points within each musical genre diversifies. For example, the points in Fig. 2d cover a wide range which means the ambiguity/variance of MSR in this group is high, however, the points in Fig. 2h are very close with each other spreading in a small range, and thus the ambiguity/variance of MSR in this group is relatively lower in contrast. The assembly of these points in Fig. 2a also shows the difference in the MSR value distribution of different musical genres, which inspires us that we can distinguish different musical genres based on their covering area of the points on the  $\alpha_{min} - \alpha_{max}$  coordinate system.

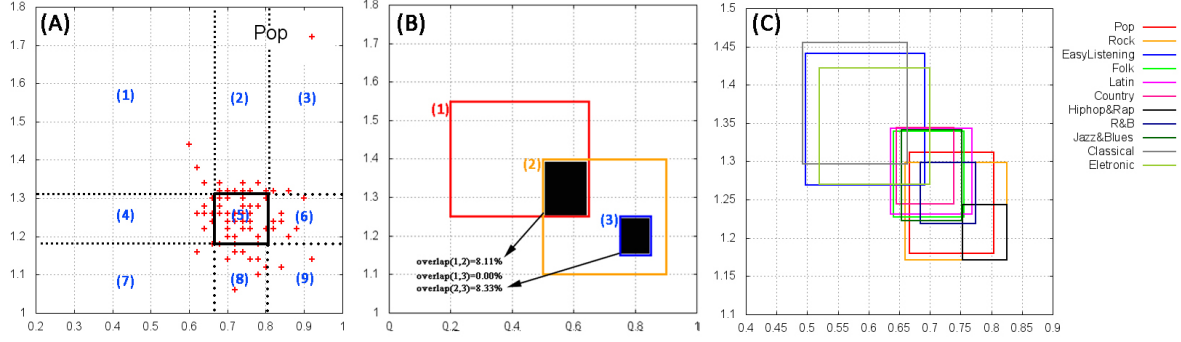
We use a rectangle to mark the edges of the MSR region (a.k.a. MSR rectangle) on the  $\alpha_{min} - \alpha_{max}$  coordinate system for a given genre. The rectangle is located by the following four coordinates:

$$(Mean(\alpha_{min}) \pm StdDev(\alpha_{min}), Mean(\alpha_{max}) \pm StdDev(\alpha_{max})), \quad (1)$$

where  $Mean(\alpha_{min})$  and  $StdDev(\alpha_{min})$  represent the mean and the standard deviation of  $\alpha_{min}$  in the given musical genre, respectively. So are  $Mean(\alpha_{max})$  and  $StdDev(\alpha_{max})$ . These values are shown in Table 1. Fig. 3a shows the MSR rectangle of Pop music. The larger the MSR rectangle is, the more ambiguity of MSR the given musical genre has. The MSR rectangle divides the  $\alpha_{min} - \alpha_{max}$  coordinate system into 9 parts on the first quadrant as illustrated in Fig. 3a. Part 9 can be considered as the ‘safe’ stretching area of the given musical genre since the points in this region are within the acceptable stretching range. Parts 1,2,3,4,7 are the ‘dangerous’ stretching areas of the given musical genre by contrast. Furthermore, Parts 5,6,8 can therefore be considered as the transition areas from the ‘dangerous’ stretching region to the ‘safe’ stretching region. If the music stretching tasks<sup>1,2,4-6</sup> are performed within the ‘safe’ region, the stretched results will be the most likely to be accepted by general audience. On the contrary, if stretched in the ‘dangerous’ region, the results will usually be unacceptable for general audience. Since different musical genres lead to different MSR rectangles, these MSR rectangles can be used to distinguish different musical genres from the perspective of MSR which has never been studied in the literature.

The area of MSR rectangle, i.e.,  $4 \times StdDev(\alpha_{min}) \times StdDev(\alpha_{max})$  can be used to measure the relative MSR ambiguity of musical genres. Moreover, the relationship of MSR rectangles from different musical genres falls into three categories: (1)





**Figure 3.** The MSR-based similarity of musical genres. (A) The MSR rectangle of *Pop* music containing points on the  $\alpha_{min} - \alpha_{max}$  coordinate system. (B) An example of the possible relationship between MSR rectangles. (C) The illustration of the MSR rectangles of all musical genres.

Genre	Pop	Rock	Easy Listening	Folk	Latin	Country	Hip-hop&Rap	R&B	Jazz&Blues	Classical	Electronic
Pop	<b>1.0</b>	0.713	0.021	0.323	0.334	0.219	0.159	<b>0.395</b>	0.346	0.0	0.032
Rock		<b>1.0</b>	0.018	0.255	0.259	0.168	<b>0.244</b>	<b>0.339</b>	0.275	1.7e-4	0.025
Easy Listening			<b>1.0</b>	0.082	0.092	0.088	0.0	0.005	0.063	0.658	0.748
Folk				<b>1.0</b>	0.808	0.675	0.002	0.344	0.822	0.024	0.113
Latin					<b>1.0</b>	0.625	0.009	0.351	0.692	0.031	0.125
Country						<b>1.0</b>	0.0	0.223	0.648	0.024	0.125
Hip-hop&Rap							<b>1.0</b>	0.042	0.0	0.0	0.0
R&B								<b>1.0</b>	0.380	0.0	0.014
Jazz&Blues									<b>1.0</b>	0.010	0.092
Classical										<b>1.0</b>	0.492
Electronic											<b>1.0</b>

**Table 2.** The MSR-based similarity between musical genres. The similarity matrix is symmetric, and only half the matrix is shown. The inclusion relationship is in **bold** mode and the exclusion relationship is in *italic* mode. The others are in the intersection relationship.

**Inclusion** — one rectangle is totally included in another rectangle, or two rectangles are exactly the same; (2) **Exclusion** — two rectangles are totally disjoint; (3) **Intersection** — two rectangles have the common overlap as well as the disjoint part. Fig. 3b shows an example of the possible relationship between musical genres based on their MSR rectangles. *Rect* 1 is intersected with *Rect* 2, while *Rect* 3 is included in *Rect* 2 and excluded from *Rect* 1. The overlaps of rectangles are filled with black, and we can compute the overlap ratio using the Jaccard similarity:

$$Sim(Rect_i, Rect_j) = \frac{Rect_i \cap Rect_j}{Rect_i \cup Rect_j}. \quad (2)$$

$Sim(Rect_i, Rect_j)$  measures how much similarity the musical genres that  $Rect_i$  and  $Rect_j$  represent, share from the perspective of MSR. The relationships of MSR rectangles of all musical genres are shown in Fig. 3c. We can see that all the inclusion, the exclusion and the intersection relationship exist among the musical genres presented in this paper.

Next, we computed the MSR-based similarity between musical genres using Eq. 2, and the results are shown in Table 2. Since the similarity matrix is symmetric, only half the matrix is shown. Obviously, the similarity between two musical genres whose MSR rectangles satisfy the *exclusion* relationship is zero (in *italic* mode). The *inclusion* relationship (in **bold** mode) is found that  $R\&B \subseteq Pop$ ,  $R\&B \subseteq Rock$ , and  $Hip-hop\&Rap \subseteq Rock$  where  $\subseteq$  is the inclusion operator. Most of the relationship observed is *intersection*. However, the ratio of overlap in the intersection relationship varies between different pairs of musical genres. The MSR-based similarity offers a new look to explore the relationship between musical genres. We can see from Table 2 that high MSR-based similarity is observed between some musical genre pairs, e.g., *Pop-Rock*, *Folk-Latin*, *Folk-Jazz&Blues*. This kind of similarity may be very difficult or even impossible to study using meta-data, audio content or cultural backgrounds since MSR-based similarity is related to human psychoacoustic perception of music as well as people's degrees of self-adaption to changes of music such as tempo, event density and lyrics density.

We believe that our work has made prospective attempts in studying the relationship between MSR and musical genres, and more issues remain to be further investigated in the future. Our findings on the MSR hold not only on digital music recordings which are stretched using complex signal processing algorithms.<sup>1,2,4-6</sup> It can also be applied in live music performance, for example, the acceleration and the deceleration of live piano or violin performances to accompany singers in a concert.

## References

1. Verhelst, W. & Roelands, M. An overlap-add technique based on waveform similarity (wsola) for high quality time-scale modification of speech. In *IEEE international conference on acoustics, speech and signal processing*, 554–557 (1993).
2. Verhelst, W. Overlap-add methods for time-scaling of speech. *Speech Communication* **30**, 207–221 (2000).
3. Chen, J. & Wang, C. Automatic music stretching resistance classification using audio features and genres. *IEEE Signal Processing Letters* **20**, 1249–1252 (2013).
4. Liu, Z., Wang, C., Wang, J., Wang, H. & Bai, Y. Adaptive music resizing with stretching, cropping and insertion. *Multimedia System* **19**, 359–380 (2013).
5. Liu, Z., Wang, C., Bai, Y., Wang, H. & Wang, J. Musiz: a generic framework for music resizing with stretching and cropping. In *ACM Multimedia*, 523–532 (2011).
6. Liu, Z., Wang, C., Guo, L., Bai, Y. & Wang, J. Lydar: a lyrics density based approach to non-homogeneous music resizing. In *IEEE international conference on multimedia and expo*, 310–315 (2010).
7. Wenner, S., Bazin, J.-C., Sorkine-Hornung, A., Kim, C. & Gross, M. Scalable music: Automatic music retargeting and synthesis. *Eurographics* **32**, 345–354 (2013).
8. Brennan, D. & Stevens, C. The effect of pitch, tempo and proportional pitch and tempo manipulation on memory of familiar musical excerpts. In *International conference on music perception and cognition*, 1771–1778 (2006).
9. Berlin, P., Zatorre, R. J., Lafaille, P., Ahad, P. & Pike, B. Voice-selective areas in human auditory cortex. *Nature* **403**, 309–312 (2000).
10. Madison, G. & Paulin, J. Ratings of speed in real music as a function of both original and manipulated beat tempo. *Journal of the Acoustical Society America* **128**, 3032–3040 (2010).
11. Bisesi, E. & Vicario, G. B. Psychoacoustic aspects of the speed of melody performance. In *International conference of students of systematic musicology*, 7–11 (2008).
12. Large, E. & Palmer, C. Perceiving temporal regularity in music. *Mathematical Behavior* **26**, 1–37 (2002).
13. Jones, M. & Boltz, M. Dynamic attending and responses to time. *Psychological Review* **96**, 459–491 (1989).
14. Drake, C., Jones, M. & Baruch, C. The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* **77**, 251–288 (2000).
15. Large, E. & Jones, M. The dynamics of attending: how people track time-varying events. *Psychological Review* **106**, 119–159 (1999).
16. Moelants, D. & Mckinney, M. F. Tempo perception and musical content: What makes a piece fast, slow or temporally ambiguous? In *International conference on music perception and cognition*, 558–562 (2004).
17. Mckinney, M. F. & Moelants, D. Deviations from the resonance theory of tempo induction. In *International conference on interdisciplinary musicology*, 124–125 (2004).
18. Lee, E., Nakra, T. M. & Borchers, J. You're the conductor: a realistic interactive conducting system for children. In *International conference on new interfaces for musical expression*, 68–73 (2004).
19. Quinn, S. & Watt, R. The perception of tempo in music. *Perception* **35**, 267–280 (2006).
20. Scaringella, N., Zoia, G. & Mlynek, D. Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine* **23**, 133–141 (2006).
21. Li, T., Ogihara, M. & Li, Q. A comparative study on content-based music genre classification. In *SIGIR*, 282–289 (2003).
22. Tzanetakis, G., Essl, G. & Cook, P. Automatic musical genre classification of audio signals. *IEEE Transaction on Speech and Audio Processing* **10**, 293–302 (2002).
23. Bagci, U. & Erzin, E. Automatic classification of musical genres using inter-genre similarity. *IEEE Signal Processing Letters* **14**, 521–524 (2007).
24. Roucos, S. & Wilgus, A. High quality time-scale modification for speech. In *IEEE international conference on acoustics, speech and signal processing*, 493–496 (1985).
25. Chen, J. & Wang, C. RESIC: A tool for music stretching resistance estimation. In *MultiMedia Modeling*, Lecture Note in Computer Science, 386–389 (2014).

## **Acknowledgements**

We would like to thank all the volunteers who participated in the listening experiments for their contributions which form the basis of this paper.